

Docket No.: 62807-172

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of	:	Customer Number: 20277
	:	
Akihisa NAGAMI, et al.	:	Confirmation Number:
	:	
Serial No.:	:	Group Art Unit:
	:	
Filed: March 11, 2004	:	Examiner:
	:	
For: TRAFFIC CONTROL APPARATUS AND SERVICE SYSTEM USING THE SAME	:	

**CLAIM OF PRIORITY AND
TRANSMITTAL OF CERTIFIED PRIORITY DOCUMENT**

Mail Stop CPD
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

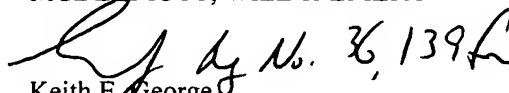
In accordance with the provisions of 35 U.S.C. 119, Applicants hereby claim the priority of:

Japanese Patent Application No. JP 2003-418905, filed on December 17, 2003.

cited in the Declaration of the present application. A certified copy is submitted herewith.

Respectfully submitted,

MCDERMOTT, WILL & EMERY


Keith E. George
Registration No. 34,111

600 13th Street, N.W.
Washington, DC 20005-3096
(202) 756-8000 KEG:gav
Facsimile: (202) 756-8087
Date: March 11, 2004

日本国特許庁 *McDermott, Will & Emery*
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: 2003年12月17日

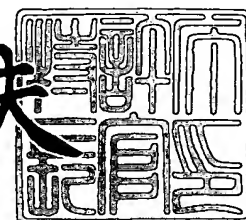
出願番号
Application Number: 特願2003-418905
[ST. 10/C]: [JP 2003-418905]

出願人
Applicant(s): 株式会社日立製作所

2004年 3月 2日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



【書類名】 特許願
【整理番号】 K03011161A
【あて先】 特許庁長官殿
【国際特許分類】 G06F 12/00
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
 システム開発研究所内
 【氏名】 永見 明久
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
 システム開発研究所内
 【氏名】 小川 祐紀雄
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
 システム開発研究所内
 【氏名】 中原 雅彦
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
 システム開発研究所内
 【氏名】 横田 大輔
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
 システム開発研究所内
 【氏名】 上岡 功司
【特許出願人】
 【識別番号】 000005108
 【氏名又は名称】 株式会社 日立製作所
【代理人】
 【識別番号】 100075096
 【弁理士】
 【氏名又は名称】 作田 康夫
【選任した代理人】
 【識別番号】 100100310
 【弁理士】
 【氏名又は名称】 井上 学
【手数料の表示】
 【予納台帳番号】 013088
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1

【書類名】 特許請求の範囲**【請求項 1】**

サーバ装置に対してサービス要求を発行する複数のクライアント装置と、前記クライアント装置からのサービス要求を受けて、サービスを提供するサーバ装置とからなるサービスシステムにおいて、前記クライアント装置と前記サーバ装置間のトラフィックを制御するトラフィック制御装置であって、

前記クライアント装置から前記サーバ装置へのサービス要求を受信する手段と、

前記サービス要求に対して前記サーバ装置から送信される応答を受信する、前記クライアント装置の受信能力に応じて、前記クライアント装置の前記サーバ装置への同時接続数を制御する手段と、

前記複数のクライアント装置から受信した前記サービス要求について、前記同時接続数に従った前記サーバ装置への中継処理を行う手段と、を備える。

【請求項 2】

請求項 1 のトラフィック制御装置は、

前記クライアント装置の前記受信能力を測定する手段を備え、

前記同時接続数を制御する手段は、前記測定結果に基づき制御する。

【請求項 3】

請求項 1 のトラフィック制御装置は、

前記サーバ装置により提供される前記応答の待ち時間を予測する手段と、

前記待ち時間が一定以上になる場合には、前記要求を拒否するアクセス規制メッセージを送信する手段を備える。

【請求項 4】

請求項 1 のトラフィック制御装置は、

前記クライアント装置のデータ受信能力に応じて、前記サーバ装置に前記要求を中継する優先度を変更する手段を備える。

【請求項 5】

請求項 1 のトラフィック制御装置において、

前記クライアント装置がサービスを受信した時間を観測して、前記クライアント装置のデータ受信能力を算出するクライアント能力測定手段を備える。

【請求項 6】

請求項 1 のトラフィック制御装置において、

前記サーバ装置が前記応答を送信した時間を観測して、前記クライアント装置のデータ受信能力を算出するクライアント能力測定手段を備える。

【請求項 7】

請求項 4 のトラフィック制御装置において、

後から受信した要求の優先度が高い場合、すでに前記クライアント装置から受信している要求に対して、アクセス規制を行う手段を備える。

【請求項 8】

請求項 1 に記載のトラフィック制御装置であって、

前記クライアント装置のデータ受信能力に応じて、前記サーバ装置に中継する前記要求の優先度を変更する手段を備える。

【請求項 9】

請求項 8 に記載のトラフィック制御装置であって、

前記クライアント装置に対する平均応答時間を一定時間以内にする手段を備える。

【請求項 10】

請求項 1 に記載のトラフィック制御装置であって、

前記サーバ装置に前記要求を転送する前に、前記クライアント装置に対して、前記要求の最大処理時間を提示する手段を備える。

【請求項 11】

クライアント装置からのサービス要求を受けてサービスを提供するサーバ装置と、前記

クライアント装置と前記サーバ装置間のトラフィックを制御するトラフィック制御装置とからなるサービスシステムであって、

前記トラフィック制御装置は、

前記クライアント装置から前記サーバ装置へのサービス要求を受信する手段と、

前記サービス要求に対して前記サーバ装置から送信される応答を受信する、前記クライアント装置の受信能力に応じて、前記クライアント装置の前記サーバ装置への同時接続数を制御する手段と、

前記複数のクライアント装置から受信した前記サービス要求について、前記同時接続数に従った前記サーバ装置への中継処理を行う手段と、を備え、

前記サーバ装置は、前記サービス要求に対する前記応答を、前記トラフィック制御装置へ送信する手段を備える。

【請求項 12】

請求項 11 に記載のサービスシステムであって、

前記トラフィック制御装置は、

前記クライアント装置のデータ受信能力に応じて、前記サーバ装置に中継する前記要求の優先度を変更する手段を備える。

【請求項 13】

請求項 10 に記載のサービスシステムであって、

前記トラフィック制御装置は、

クライアントに対する平均応答時間を一定時間以内にする手段を備える。

【請求項 14】

請求項 10 に記載のサービスシステムであって、

前記トラフィック制御装置は、前記サーバ装置に前記要求を転送する前に、前記クライアント装置に対して、前記要求の最大処理時間を提示する手段を備える。

【書類名】 明細書**【発明の名称】** トラフィック制御装置およびそれを用いたサービスシステム**【技術分野】****【0001】**

本発明は、クライアント装置とサーバ装置データ通信中継システムを用いたクライアント装置とサーバ装置間のデータ通信方法と、それを用いたサーバアクセスのサービスに関する。

【背景技術】**【0002】**

近年のインターネットの急速な普及に伴い、インターネットへのアクセス方法が多様化してきている。

【0003】

アクセス網のブロードバンド化を背景に、広帯域、低遅延なネットワークを介した高性能なPCからのサーバアクセスが増加する一方、携帯電話によるWebアクセスやPHS網を利用したPDAなどからのWebアクセスなど、遅延が大きく、帯域の狭いネットワークを介した、性能の低い機器からのアクセスも増加している。

【0004】

従来、WebサーバやFTPサーバでは、サーバ装置(以下、サーバという)にアクセスが集中して、サービスがダウンしたり性能が低下したりすることを防ぐため、クライアント装置(以下、クライアントという)との同時接続数を制限し、運用していた(例えば、特許文献1参照)。

【0005】

このWebサーバへのアクセスで利用される通信プロトコルであるHTTP (Hypertext Transfer Protocol) の詳細および使い方については、非特許文献1に詳しく説明されている。

【0006】

【特許文献1】 米国特許出願公開第2003/0028616号明細書

【0007】

【非特許文献1】 R. Fielding, UC Irvine, J. Gettys, Compaq/W3C, J. Mogul, Compaq, H. Frystyk, W3C/MIT, L. Masinter, Xerox, P. Leach, Microsoft, T. Berners-Lee, W3C/MIT著, "Hypertext Transfer Protocol -- HTTP/1.1", [online], 1999年6月, IETF, インターネット<URL:<http://www.ietf.org/rfc/rfc2616.txt>>

【発明の開示】**【発明が解決しようとする課題】****【0008】**

従来のインターネット環境では、クライアントのネットワーク環境に大差はなく、サーバがダウンすることを防ぐためにはクライアントとの同時接続数を制限するだけで十分であった。しかし、上述のようなインターネット環境の変化により、次のような問題が発生している。

【0009】

PDAや携帯電話など帯域の狭いネットワークを介したクライアントからのアクセスが集中した場合には、一つ一つのクライアントに対する負荷は高くないが、サーバでの処理時間が長くなるため、サーバの最大同時接続数は大きく設定しないと性能を十分に活用することができない。一方、光ファイバー網に接続している高性能なPCからのアクセスが集中した場合には、ネットワーク遅延も、クライアントの受信遅延も小さいため、一つ一つのクライアントに対する処理負荷が高く、サーバでの同時接続数を大きく設定しているとサーバが停止してしまうことがある。このように、クライアントの性能差が広がったことで、クライアントとの同時接続数の制限だけではサーバの停止を防止することは難しくなっている。

【0010】

また、サーバに接続してサービス要求を送信してからサービスを受けるまでの時間もサ



サーバの負荷状態によりさまざまであり、サーバに接続したものの、なかなかサービスを受けられないといった状況が発生していた。

【0011】

このため、サーバによるサービス提供を、レベルを下げることなく、より安定して行える技術が求められている。

【課題を解決するための手段】

【0012】

本発明は、クライアントのネットワーク特性を考慮してサーバとの同時接続数を制御し、サーバを安定して稼働させる技術を提供する。

【0013】

また、サービス要求に対する応答時間を推定し、サービスを提供するためにある一定時間以上の時間が必要と予測する要求に関しては、アクセス規制メッセージを即座に応答し、利用者に不確定な長時間、サービスを待たせない技術を提供する。

【0014】

具体的な本発明の一態様においては、まず、クライアントからのサーバへのアクセスにおいて、データ処理をしながら中継処理を行うトラフィック制御装置をサーバとクライアントの間に設ける。この構成において、クライアントからサーバへアクセスする際、トラフィック制御装置が、クライアントからの要求を受け付け、サーバへ要求を転送し、サーバはクライアントから要求されたサービスを提供する。

【0015】

本発明によるトラフィック制御装置の特徴的な動作は以下のとおり。

【0016】

すなわち、トラフィック制御装置は、サーバからの応答をクライアントに転送する際、クライアントのデータ受信能力を推定しつつ応答の転送を行う。

【0017】

また、トラフィック制御装置は、クライアントからの要求を受け付ける際、要求された宛先に対するサービスを提供できるまでにかかる時間を推定し、ある一定以上時間が必要な場合には、アクセス拒否を行う。これにより、クライアントは不確定な長時間、サービス提供を待たなければならないということがなくなる。

【0018】

また、トラフィック制御装置は、クライアントからの要求をキューイングするための手段を持ち、クライアントから受け付けた要求のサーバへの転送タイミングを制御することができる。

【0019】

また、トラフィック制御装置は、クライアントからの要求をサーバに転送する際、サーバのデータ送信能力とクライアントのデータ受信能力に応じてサーバに対する同時接続数を制御する。

【発明の効果】

【0020】

本発明によれば、過剰に要求が到達してサーバが停止することなく、また、アクセス制限によりスループットが低下することなくなり、クライアントに対するサービスレベルが低下することを防ぐことができる。これにより、サーバに対する投資を抑えることができるとともに、安定性を向上することができる。

【0021】

また、サーバの提供するサービスの処理スループットを向上させることができる。

【0022】

さらに、要求を受け付けたクライアントに対しては、不確定な長時間待たせる可能性を低くすることができる。

【発明を実施するための最良の形態】

【0023】

図1は、本実施例のデータ通信装置を使ったシステムの構成図である。

【0024】

本実施例では、各クライアント装置(1)と各サーバ装置(2)は1台以上のトラフィック制御装置(3)を介して通信路(4)で接続した構成とする。トラフィック制御装置(3)は、クライアント装置(1)とサーバ装置(2)間のデータ通信を中継する。すなわち、クライアント装置(1)からサーバ装置(2)へのサービス要求(50)は、必ず、トラフィック制御装置(3)を介して送られ、サーバ装置(2)からクライアント装置(1)への応答(60)も、トラフィック制御装置(3)を介して行われる。なお、ここでいう通信路(4)とは、必ずしも物理的な通信回線に一致する必要はなく、物理的な通信回線上で実現された論理的な通信経路であってよい。

【0025】

図2は、本実施例に関わるクライアント装置(1)、サーバ装置(2)、トラフィック制御装置(3)、各装置の物理的な一構成例を示す。これらの装置は、図2に示すとおり、物理的には一般的な情報処理装置でよい。具体的には、各装置情報処理装置は、例えば、プロセッサ(101)、メモリ(102)、外部記憶装置(103)、通信装置(104)、オペレータ入出力装置(105)がバスなどの内部通信線(106)で接続した形で構成する。

【0026】

各装置のプロセッサ(101)は、メモリ(102)に格納された情報処理プログラム(108)を実行することにより、以下の実施例において説明する処理を実現する。

【0027】

メモリ(102)は、情報処理プログラム(108)以外にも、情報処理プログラム(108)から参照される各種データも保持する。

【0028】

外部記憶装置(103)は、情報処理プログラム(108)や各種データを不揮発な形で保持する。プロセッサ(101)は情報処理プログラム(108)を実行することにより、外部記憶装置に指示して、必要なプログラムやデータをメモリ(102)にロードしたり、逆に、メモリ(102)上の情報処理プログラム(108)やデータを外部記憶装置(103)に格納したりする。情報処理プログラム(108)は、予め外部記憶装置(103)に格納されていても良い。または必要に応じて、情報処理装置が利用可能な、可搬型記憶媒体や、通信媒体すなわち通信回線または通信回線を伝搬する搬送波、を介して、外部の装置から導入されても良い。

【0029】

通信装置(104)は、通信回線(107)に接続し、情報処理プログラム(108)の指示により、他の情報処理装置や通信機器にデータを送信したり、逆に、他の情報処理装置や通信機器から、データを受信して、メモリ(102)に格納する。装置間の論理的な通信路(4)は、本通信装置を介し、物理的な通信回線(107)を通して実現される。

【0030】

オペレータ入出力装置(105)は、文字通り、オペレータとの間でデータ入出力を行う。

【0031】

内部通信線(106)は、プロセッサ(101)、メモリ(102)、外部記憶装置(103)、通信装置(104)、オペレータ入出力装置(105)が互いに通信を行うためのものであって、たとえばバスで構成する。

【0032】

クライアント装置(1)、サーバ装置(2)、データ通信中継システム(3)は、必ずしも物理的に異なった構成にする必要はなく、それぞれの機能の違いは、各装置上で実行する情報処理プログラム(108)で実現する形でよい。

【0033】

以下では、実施例を説明する際、処理部という言葉で、本実施例中の構成要素を説明するが、各処理部は、論理的な構成を示すものであり、物理的装置で実現しても、情報処理プログラム（108）が実行されることにより実現された一機能であってもいずれでもよい。また、クライアント装置（1）、サーバ装置（2）、トラフィック制御装置（3）も、それぞれ独立した物理的な装置である必要はなく、1つの装置において実現されてもよい。さらに、それぞれの各処理部が1つの装置で構成されている必要は無く、異なる装置に分散して実現されていてもよい。

【0034】

図3は本実施例のトラフィック制御装置（3）の一構成図である。

【0035】

本実施例のトラフィック制御装置（3）は、クライアント装置（1）からの要求（50）の受信を行う要求受信部（21）、要求（50）をサーバ装置（2）へ送信する要求送信部（22）、クライアント装置（1）のアドレスからクライアント能力を決定するクライアント能力ランク部（23）、クライアント装置（1）からの要求をキューに入れて管理するアクセス管理部（24）、サーバ装置（2）からの応答（60）を受信する応答受信部（25）、クライアント装置（1）へ応答（60）を送信する応答送信部（27）、クライアント装置（1）のデータ受信能力を測定するクライアント能力測定部（26）と、クライアント装置（1）からの要求を管理するキュー管理表（31）、サーバ装置（2）に対するアクセス状況を管理するアクセス管理表（32）、クライアント装置（1）の受信能力を管理するクライアントデータ受信能力表（33）を備える。

【0036】

さらに、サーバ装置（2）側のネットワークを観測し、応答（60）が送信された時刻を測定するサーバ側クライアント能力測定部（28）とクライアント装置（1）側のネットワークを観測し、応答（60）が受信された時刻を測定するクライアント側クライアント能力測定部（29）を備えてもよい。

【0037】

図11は、本実施例の要求キュー管理表（31）の一構成例を示す。要求キュー管理表（31）は、宛先フィールド（3102）をキーとする表で、同宛先フィールド（3102）で示された特定の宛先への要求のキュー管理のために使用する。

【0038】

本実施例の要求キュー管理表（31）の各エントリ（3101）は、宛先フィールド（3102）の他、キューイングしている要求の数を表すキュー中の要求数フィールド（3103）、宛先フィールド（3102）のサービスの最大の待ち時間を表す最大待ち時間フィールド（3104）、応答時間の平均を表す応答時間平均フィールド（3105）、応答時間の平均を算出するために必要な応答時間総和フィールドと（3106）応答時間処理数フィールド（3107）と、要求（50）をリンクリストで管理する要求リンクリストフィールド（3108）を含む。宛先フィールド（3102）と、最大待ち時間フィールド（3104）は、予めオペレータが予め設定する。

【0039】

図12は、本実施例のアクセス管理表（32）の一構成例を示す。アクセス管理表（32）は、宛先フィールド（3202）をキーとする表で、同宛先フィールド（3202）で示された特定の宛先へのアクセス状況を管理するために使用する。

【0040】

本実施例のアクセス管理表（32）の各エントリ（3201）は、宛先フィールド（3202）の他、その宛先サーバの能力を表す宛先サーバ能力フィールド（3203）、現在接続中のクライアントの転送能力を表すクライアント能力の和フィールド（3204）、サーバに対して同時に接続できるクライアントの許容最大数を表す最大接続数フィールド（3205）、現在接続中のクライアントの数を表す現在の接続数フィールド（3206）を含む。宛先フィールド（3202）と、宛先サーバ能力フィールド（3203）と、最大接続数フィールド（3205）は、予めオペレータが設定する。

【0041】

図13は、本実施例のクライアントデータ受信能力表(33)の一構成例を示す。クライアントデータ受信能力表(33)は、クライアントのアドレス(3302)をキーとする表で、同クライアントアドレスフィールド(3302)で示された特定のクライアントのデータ受信能力を管理するために使用する。

【0042】

本実施例のクライアントデータ受信能力表(33)の各エントリ(3301)は、クライアントアドレスフィールド(3302)の他、クライアントのデータ受信能力をあらわすクライアントデータ受信能力フィールド(3303)、データ受信能力を算出するために用いるデータ送信開始時刻、データ受信完了時刻、データサイズを表す送信開始時刻フィールド(3304)、受信完了時刻フィールド(3305)、データサイズフィールド(3306)を含む。

【0043】

図14は、宛先サーバアドレス(51)と要求サービス名(52)とからなる、要求(50)の一構成例を示す。

【0044】

図4は、本実施例のトラフィック制御装置(3)が、クライアント装置が送信した要求(50)をサーバ装置(2)に転送する処理フローを示す。

【0045】

クライアント装置(1)が、トラフィック制御装置(3)経由で宛先サーバ装置(2)へのサービス要求を送信する。(処理1001)このとき、クライアント装置(1)自身がトラフィック制御装置(3)を意識して、トラフィック制御装置(3)宛で送信してもよいし、クライアント装置(1)は意識せず、データ通信を中継するルータ等が、宛先サーバ装置(2)宛の要求(50)をトラフィック制御装置(3)に転送してもよい。

【0046】

トラフィック制御装置(3)の要求受信部(21)は、クライアント装置(1)からのサービス要求(50)を受け付け受信する(処理1002)。

【0047】

要求受信部(21)は、受信した要求を解析し、クライアントアドレス、宛先サーバアドレス(51)、要求サービス名(52)を特定する(処理1003)。

【0048】

クライアント能力ランク部(23)は、クライアントアドレスをキーにクライアントデータ受信能力表(33)を検索する。エントリがあればそのエントリのクライアントデータ受信能力フィールド(3303)の値を要求受信部(21)に返す。対応するエントリがない場合には、各フィールドはデフォルト値と同じ物を用いて新しくエントリを作成し追加する(処理1004)。

【0049】

要求受信部(21)は、要求(50)をクライアントデータ受信能力(3303)とともにアクセス管理部(24)に送信する(処理1005)。

【0050】

アクセス管理部(24)は、受信した要求(50)に対応するエントリを求めてアクセス管理表(32)を検索する。対応するエントリがない場合には、各フィールドはデフォルト値と同じ物を用いて新しくエントリを作成し追加する(処理1006)。

【0051】

アクセス管理部(24)は、処理1006で得たエントリの情報を元に宛先サーバ装置(2)にアクセス可能か判定する(処理1007)。

【0052】

処理1007の結果アクセス可能な場合には、アクセス管理部(24)は、対応するアクセス管理エントリ(3201)のクライアント能力の和フィールド(3204)の値に現在処理中のクライアントの能力値を加算する。また、現在の接続数フィールド(320

6) の値を 1 増加する。その後、要求 (50) を要求送信部 (22) に送信する (処理 1011)。

【0053】

要求送信部 (22) は、受信した要求 (50) をサーバ装置 (2) に送信する (処理 1012)。

【0054】

処理 1012 の後で、クライアントに対して処理時間を提示するため、クライアントに対して HTTP の chunk 形式などで応答のコンテンツの先頭として宛先の最大待ち時間フィールド (3104) の値を返信してもよい。ただし、この処理は宛先が HTML などのテキストコンテンツに限る。

【0055】

サーバ装置 (2) はサービス要求 (50) を受信し、サービスのための処理を開始する (処理 1013)。

【0056】

処理 1007 の結果アクセス不可の場合には、アクセス管理部 (24) は、要求キュー管理表 (31) を参照し、キューイング可能か判定する (処理 1008)。

【0057】

処理 1008 の結果キューイング可能な場合には、要求 (50) に対応する宛先のエントリの要求リンクリストフィールド (3108) に要求を追加し、キューの中の要求数フィールドの値を 1 増加する (処理 1010)。

【0058】

処理 1008 の結果キューイング不可の場合には、クライアント装置 (2) に対してアクセス規制のメッセージを応答する (処理 1009)。

【0059】

図 5 と図 6 は、本実施例のトラフィック制御装置 (3) がサーバ装置 (2) が送信した応答 (60) をクライアント装置 (1) に転送する処理フローを示す図である。

【0060】

図 5 において、サーバ装置 (2) は要求されたサービスに対する処理を終え、サービスをクライアント装置 (1) に提供するため、応答 (60) を送信する (処理 1101)。

【0061】

トラフィック制御装置 (3) の応答受信部 (25) は、サーバ装置 (2) が送信する応答 (60) を受信する (処理 1102)。

【0062】

クライアント能力測定部 (26) は、サーバ装置 (2) が応答 (60) の送信を開始した時刻を取得する。この取得の方法は、応答受信部 (25) が応答 (60) を受信した時間により取得する。ただし、サーバ側クライアント能力測定部 (28) が、トラフィック制御装置 (3) とサーバ装置 (2) との通信路 (4) を観測している場合には、サーバ側クライアント能力測定部 (28) がサーバ装置 (2) が応答 (60) を送信し始めた時刻を決定し、クライアント能力測定部 (26) に通知してもよい。クライアント能力測定部 (26) は、応答 (60) の送信開始時刻を取得し、クライアントデータ受信能力表 (33) のクライアントアドレスに対応するエントリ (3301) の送信開始時刻フィールド (3304) の値が設定されていなければ、取得した値を設定する (処理 1103)。送信開始時刻フィールド (3304) は、例えば、応答 (60) の先頭をクライアントに送信したときに設定するため、先頭を送信するまでは設定されていない。

【0063】

応答受信部 (25) は、応答 (60) を応答送信部 (27) に送信する (処理 1104)。

【0064】

応答送信部 (27) は、応答 (60) をクライアント装置 (1) に送信する (処理 1105)。

【0065】

クライアント装置 (1) は、応答 (60) を受信し、要求したサービスをうける (処理 1106)。

【0066】

クライアント能力測定部 (26) は、クライアント装置 (1) が応答 (60) の受信を完了した時刻を取得する。この取得の方法は、応答送信部 (27) が応答 (60) を送信し終わった時刻により取得する。ただし、クライアント側クライアント能力測定部 (29) が、トラフィック制御装置 (3) とクライアント装置 (1) との通信路 (4) を観測している場合には、クライアント側クライアント能力測定部 (29) がクライアント装置 (1) が応答 (60) を受信し終えた時刻を決定し、クライアント能力測定部 (26) に通知してもよい。クライアント能力測定部 (26) は、応答 (60) の受信完了時刻を取得し、クライアントデータ受信能力表 (33) のクライアントアドレスに対応するエントリ (3301) の受信完了時刻フィールド (3305) の値が設定されていなければ、取得した値を設定する。受信完了時刻フィールド (3305) は、例えば、応答 (60) の最後をクライアントに送信し終えたときに設定するため、最後を送信するまでは設定されていない。

さらに、転送した応答 (60) のデータサイズを同エントリのデータサイズフィールド (3306) の値に設定する (処理 1107)。

【0067】

クライアント能力測定部 (26) は、現在処理中のクライアント装置 (1) のデータ受信能力を計算するため、対応するエントリの受信完了時刻フィールド (3305) の値から送信開始時刻フィールド (3304) の値を引き、データサイズフィールド (3306) の値を割り、これをクライアントデータ受信能力フィールド (3303) の値に設定する。この計算が終了したら、送信開始時刻フィールド (3304)、受信完了時刻フィールド (3305) とデータサイズフィールド (3306) の値を消去する。ここで、クライアント能力測定部 (26) は、アクセス管理表 (32) のクライアント能力の和フィールド (3204) の値を更新してもよい (処理 1108)。

【0068】

処理 1107 と処理 1108 は、処理 1105 のクライアントに対して応答を送信している最中に、オペレータにより設定された時間間隔毎、もしくは設定された送信データサイズ毎に実行してもよい。これにより、クライアント能力の動的な変更が可能になり、ストリーミングのような、セッションが長時間にわたる場合に、そのセッションの途中でクライアント能力を動的に決定することができる。

【0069】

図 6 において、応答送信部 (27) は、アクセス管理部 (24) に送信処理が完了したことを通知する (処理 1201)。

【0070】

アクセス管理部 (24) は、応答送信部 (27) からの通知を受け、アクセス管理表 (32) から応答 (60) の送信が完了した要求 (50) を削除する (処理 1202)。

【0071】

アクセス管理部 (24) は、要求キュー管理表 (31) の要求リンクリストフィールド (3108) から該当する要求 (50) を削除し、キュー中の要求数フィールド (3103) の値を 1 減らす。次に、応答時間平均フィールド (3105) を更新するため、応答時間処理数フィールド (3107) を 1 増加し、応答時間総和フィールド (3106) に、応答時間を足し、応答時間処理数 (3107) フィールドの値で割り、応答時間平均フィールド (3105) に設定する (処理 1203)。

【0072】

アクセス管理部 (24) は、要求キュー管理表 (31) から転送できる次の要求 (50) があるかどうかを検索する (処理 1204)。

【0073】

処理 1204 の結果、転送できる次の要求 (50) がなければ、そこで処理を終了する (処理 1205)。転送できる次の要求 (50) があれば、アクセス管理部 (24) は要求キュー管理表 (31) の要求リンクリストフィールド (3108) から転送する次の要求 (50) を削除し、キュー中の要求数フィールド (3103) の値を 1 減らす。そして、削除した次の要求 (50) を要求送信部 (22) に送信する (処理 1206)。

【0074】

要求送信部 (22) は、次の要求 (50) をサーバ装置 (2) に送信する (処理 1207)。

【0075】

サーバ装置 (2) は、トラフィック制御装置 (3) から送信された次の要求 (50) を受信し、サービス提供のための処理を開始する (処理 1208)

図 7 は、処理 1004 の一実施例を表す処理フロー図である。

【0076】

クライアント能力ランク部 (23) は、要求 (50) のクライアントアドレスに対応したエントリを求めて、クライアントデータ受信能力表 (33) を検索する (処理 1501)。

【0077】

対応するエントリがあった場合には、クライアントアドレスに対応したエントリ (3301) のデータ受信能力フィールド (3303) の値を返す (処理 1502)。

【0078】

対応するエントリがなかった場合には、クライアントデータ受信能力表 (33) に新しいエントリを追加し、各フィールドはデフォルトの値を設定する。そして、デフォルトのエントリのデータ受信能力フィールド (3303) の値を返す (処理 1503)。

【0079】

図 8 は、処理 1006 の一実施例を表す処理フロー図である。

【0080】

アクセス管理部 (24) は、要求 (50) の宛先に対応したエントリを求めてアクセス管理表 (32) を検索する (処理 1511)。

【0081】

対応するエントリがなければ、アクセス管理表 (32) に現在の宛先に対する新しいエントリを作成し、各フィールドにデフォルト値を設定する (処理 1512)。

【0082】

その後、対応するエントリから宛先サーバの能力 (3203) と現在接続中のクライアント能力の和 (3204) と最大接続数 (3205) と現在の接続数 (3206) を取得する (処理 1513)。

【0083】

処理 1511 の結果、対応するエントリがあれば、処理 1513 を行い、処理 1006 を終了する。

【0084】

図 9 は、処理 1007 の一実施例を表す処理フロー図である。

【0085】

アクセス管理部 (24) は、宛先サーバ能力 (3203) がクライアント能力の和 (3204) に現在要求 (50) を処理しているクライアント装置 (1) の能力を加えた値よりも大きい確認する (処理 1521)。

【0086】

大きくなければ、アクセス不可を返し、処理 1007 を終了する。

【0087】

大きければ、設定されている最大接続数 (3205) が現在の接続数 (3206) と 1 の和よりも大きい確認する (処理 1522)。

【0088】

大きくなければ、アクセス不可を返し、大きければ、アクセス可を返して処理 1007 を終了する。

【0089】

図 10 は、処理 1008 の一実施例を表すフロー図である。

【0090】

アクセス管理部 (24) は、キュー管理表 (31) を参照し、現在の要求の宛先に対応するエントリを検索する (処理 1531)。

【0091】

処理 1531 の結果、対応するエントリがなければ、対応するエントリを新規に作成し、各フィールドにデフォルト値を設定する (処理 1532)。その後、アクセス可能を返し、処理 1008 を終了する。

【0092】

処理 1531 の結果、対応するエントリがあれば、応答時間平均フィールド (3105) の値と現在キューに格納されている要求の数を表すキュー中の要求数フィールド (3103) の値の積が、設定されている最大待ち時間フィールド (3104) をこえるか確認する (処理 1533)。つまり、要求のリンクリスト (3108) で管理している要求の最大個数 (キューの最大長) を、宛先の平均応答時間によって制御することができ、処理の待ち時間が最大待ち時間 (3104) を超えそうな場合にはキューに要求を入れない、という制御が可能である。

【0093】

最大待ち時間を超える場合は、アクセス不可を返し、最大待ち時間を超えない場合は、アクセス可を返して処理 1008 を終了する。

【0094】

以上述べたとおり、本実施例によれば、

クライアント装置 (1) との通信帯域や伝送遅延が多種多様な状況においても、サーバ装置 (2) に対するアクセス数が適切に制御されるため、過剰に要求が到達してサーバ装置 (2) がサービス提供が出来なくなることはなく、また、過剰なアクセス制限によりスループットが低下することなくなり、クライアント装置 (1) に対するサービスレベルが低下することを防ぐことができる。

【0095】

これにより、サーバ装置の性能を十分に発揮することができるため、少ないサーバ装置台数で多くのアクセスに対応することができる。つまり、サーバ装置に対する投資を抑えることができるとともに、サービス提供の安定性を向上することができる。

【0096】

また、クライアント装置は、受け付けられるか否かにかかわらず、何らかの回答が帰ってくるため、不確定な長時間待たされる可能性が低くなる。さらに、要求が受け付けられ、最大待ち時間フィールド (3104) の値が応答されれば、ある一定時間以下でサービスが提供されることを予想できるようになる。そのため、クライアント装置がサービス要求を何度も実行してサーバの負荷を高めることがなくなる。

【0097】

従来、サービス提供事業者の見積もり以上にアクセスが集中した場合、サーバ装置に過度に負荷がかかり、サービス提供が不能になるような状況が生じていたが、本実施例を適用することにより、サーバ装置による安定したサービス提供が可能になり、かつアクセス拒否を通知したユーザに対してもサービス提供の機会があるため、サーバ装置の性能を必要以上に高く見積もる必要がなくなる。

【0098】

次に、クライアント装置 (1) のネットワーク特性などのデータ受信能力に応じて優先度を変更する第二の実施例を説明する。

【0099】

本実施例では、要求キュー管理表 (31) を図 15 のように構成する。

【0100】

図11の要求のリンクリストフィールド(3108)の代わりに、優先度付き要求のリンクリスト(3110)とクライアントのデータ受信能力がこの値以上ならば優先キューに挿入する基準となる優先しきい値フィールド(3109)を持つ。優先度付き要求のリンクリスト(3110)は、優先して処理を行う優先キュー(3111)と優先しない一般キュー(3112)とからなる。

【0101】

図4の処理1010で、アクセス管理部(24)が要求キューを要求キュー管理表(31)に登録する際、クライアント装置(1)のデータ受信能力が、優先しきい値フィールド(3109)の値よりも大きければ優先キュー(3111)に追加し、大きくなければ、一般キュー(3112)に追加する。

【0102】

なお、優先キュー(3111)に追加したときに、アクセス管理表(32)のクライアント能力の和フィールド(3204)の値が宛先サーバ能力フィールド(3203)の値を超えている場合には、一般キュー(3112)のリンクリストから、既に追加した要求(50)を新しい順に削除し、クライアント能力の和フィールド(3204)の値が宛先サーバ能力フィールド(3203)の値より小さくなるようにしてもよい。さらに、削除することにより、アクセス規制のメッセージを、処理1009と同様に応答すればよい。

【0103】

さらに、図6の処理1204で、転送できる要求があるか検索する際、優先キュー(3111)を先に検索する。優先キュー(3111)に転送できる要求(50)がなければ一般キュー(3112)を検索する。

【0104】

クライアント装置のネットワーク特性によって処理優先度を変える本実施例によれば、ネットワーク帯域が広いクライアント装置に対する処理を優先して行い、応答の悪いクライアント装置に対する処理を後回しにすることができ、サーバ装置はサービスのスループットをさらに向上することができる。

【図面の簡単な説明】**【0105】**

【図1】本実施例のトラフィック制御装置を使ったシステム構成図。

【図2】本実施例に関わるクライアント装置、サーバ装置、トラフィック制御装置の物理構成図。

【図3】本実施例のトラフィック制御装置の構成図。

【図4】本実施例のトラフィック制御装置での中継処理の流れを示す処理フロー図(その1)。

【図5】本実施例のトラフィック制御装置での中継処理の流れを示す処理フロー図(その2)。

【図6】本実施例のトラフィック制御装置での通知処理の流れを示す処理フロー図(その3)。

【図7】本実施例の処理に関わる図4の処理1004の処理フロー図。

【図8】本実施例の処理に関わる図4の処理1006の処理フロー図。

【図9】本実施例の処理に関わる図4の処理1007の処理フロー図。

【図10】本実施例の処理に関わる図4の処理1008の処理フロー図。

【図11】本実施例の要求キュー管理表(31)の構成図。

【図12】本実施例のアクセス管理表(32)の構成図。

【図13】本実施例のクライアントデータ受信能力表(33)の構成図。

【図14】本実施例の要求(50)の構成図。

【図15】第二の実施例の要求キュー管理表(31)の構成図。

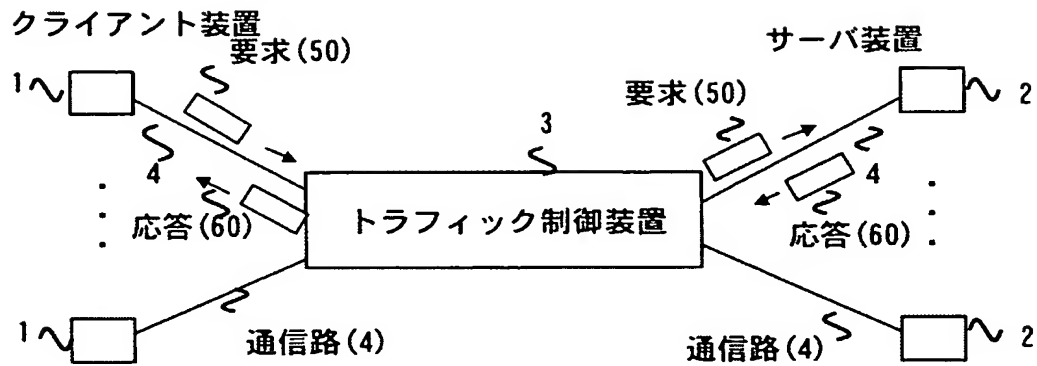
【符号の説明】**【0106】**

(1) : クライアント装置、(2) : サーバ装置、(3) : トラフィック制御装置、(4) : 通信路、(21) : 要求受信部、(22) : 要求送信部、(23) : クライアント能力ランク部、(24) : 要求キューイング部、(25) : 応答受信部、(26) : クライアント能力測定部、(27) : 応答送信部、(28) : サーバ側クライアント能力測定部、(29) : クライアント側クライアント能力測定部、(31) : キュー管理表、(32) : アクセス管理表、(33) : クライアント受信能力ランク表、(50) : 要求、(51) : 宛先サーバアドレス、(52) : 要求サービス名、(60) : 応答、(101) : プロセッサ、(102) : メモリ、(103) : 外部記憶装置、(104) : 通信装置、(105) : オペレータ入出力装置、(106) : 内部通信線、(107) : 通信回線、(108) : 情報処理プログラム、(3101) : 要求キュー管理エントリ、(3102) : 宛先フィールド、(3103) : キュー中の要求数フィールド、(3104) : 最大待ち時間フィールド、(3105) : 応答時間平均フィールド、(3106) : 応答時間総和フィールド、(3107) : 応答時間処理数フィールド、(3108) : 要求のリンクリストフィールド、(3109) : 優先しきい値フィールド、(3110) : 優先度付き要求のリンクリストフィールド、(3111) : 優先キュー、(3112) : 一般キュー、(3201) : アクセス管理エントリ、(3202) : 宛先フィールド、(3203) : 宛先サーバ能力フィールド、(3204) : クライアント能力の和フィールド、(3205) : 最大接続数フィールド、(3206) : 現在の接続数フィールド、(3301) : クライアントデータ受信能力エントリ、(3302) : クライアントアドレスフィールド、(3303) : クライアントデータ受信能力フィールド、(3304) : 送信開始時刻フィールド、(3305) : 受信完了時刻フィールド、(3306) : データサイズフィールド。

【書類名】 図面

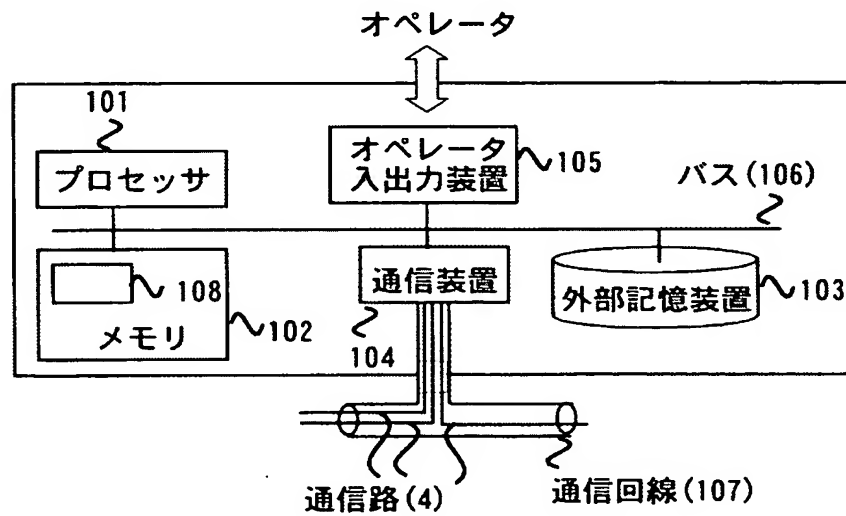
【図 1】

図 1



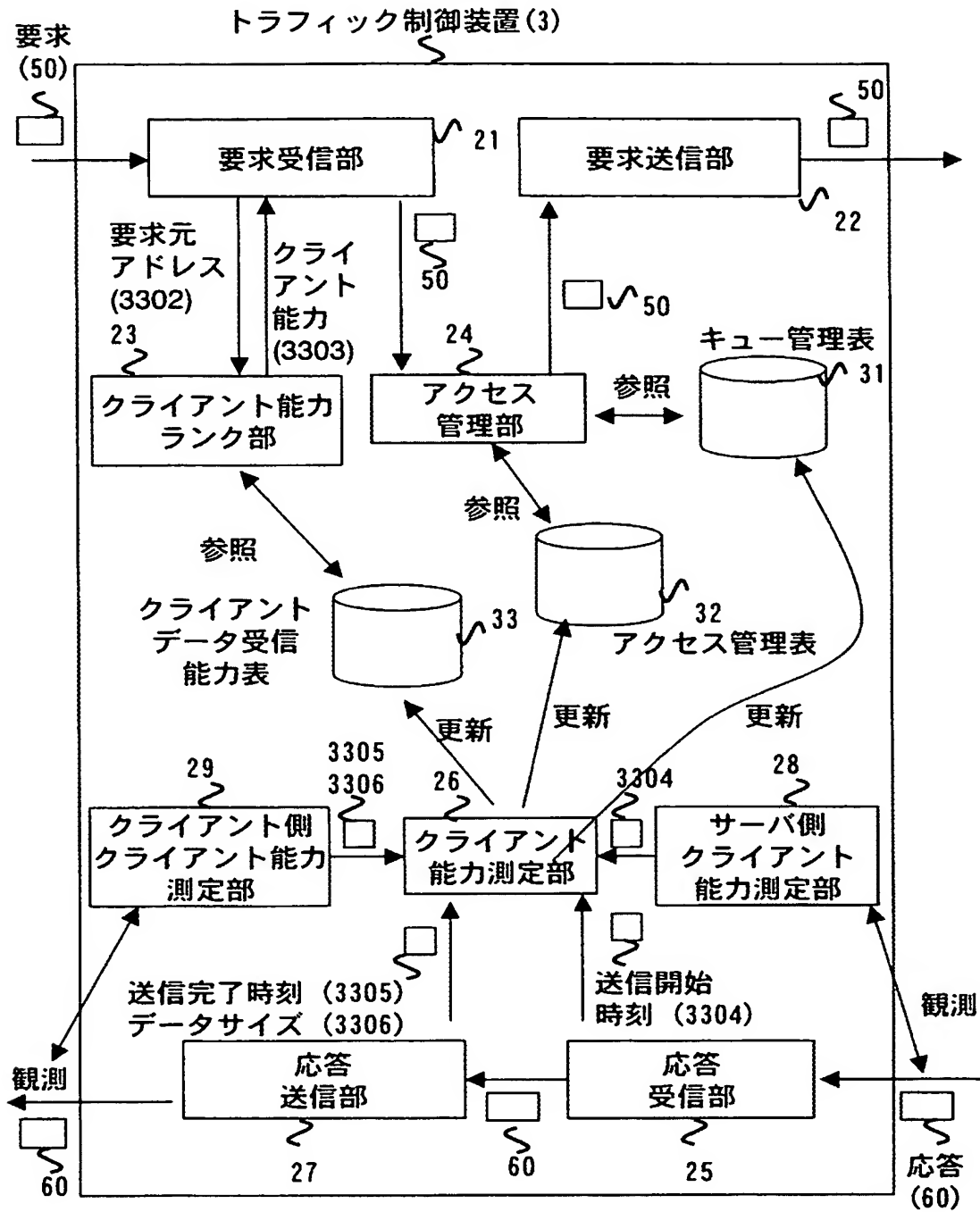
【図 2】

図 2



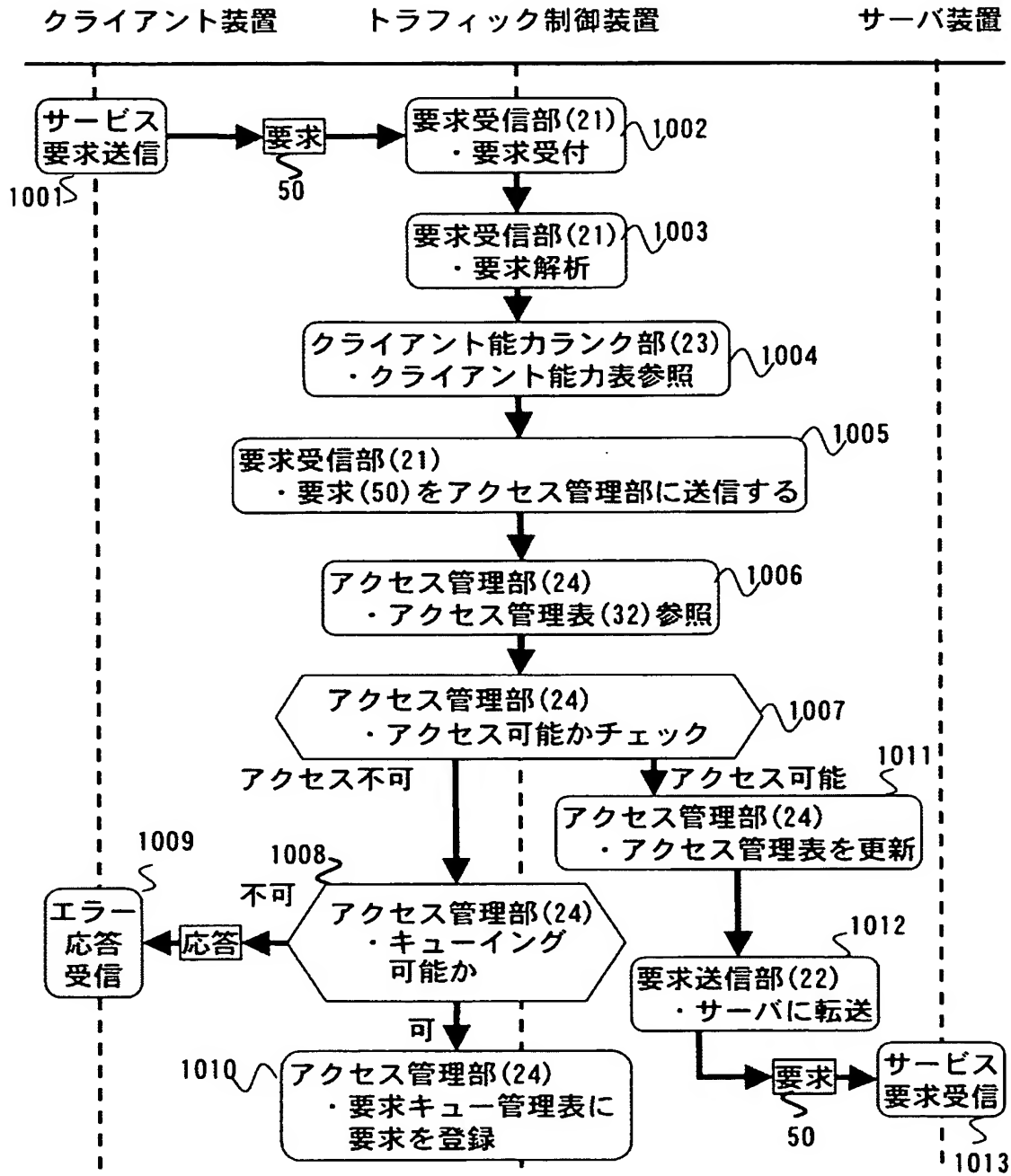
【図 3】

図 3

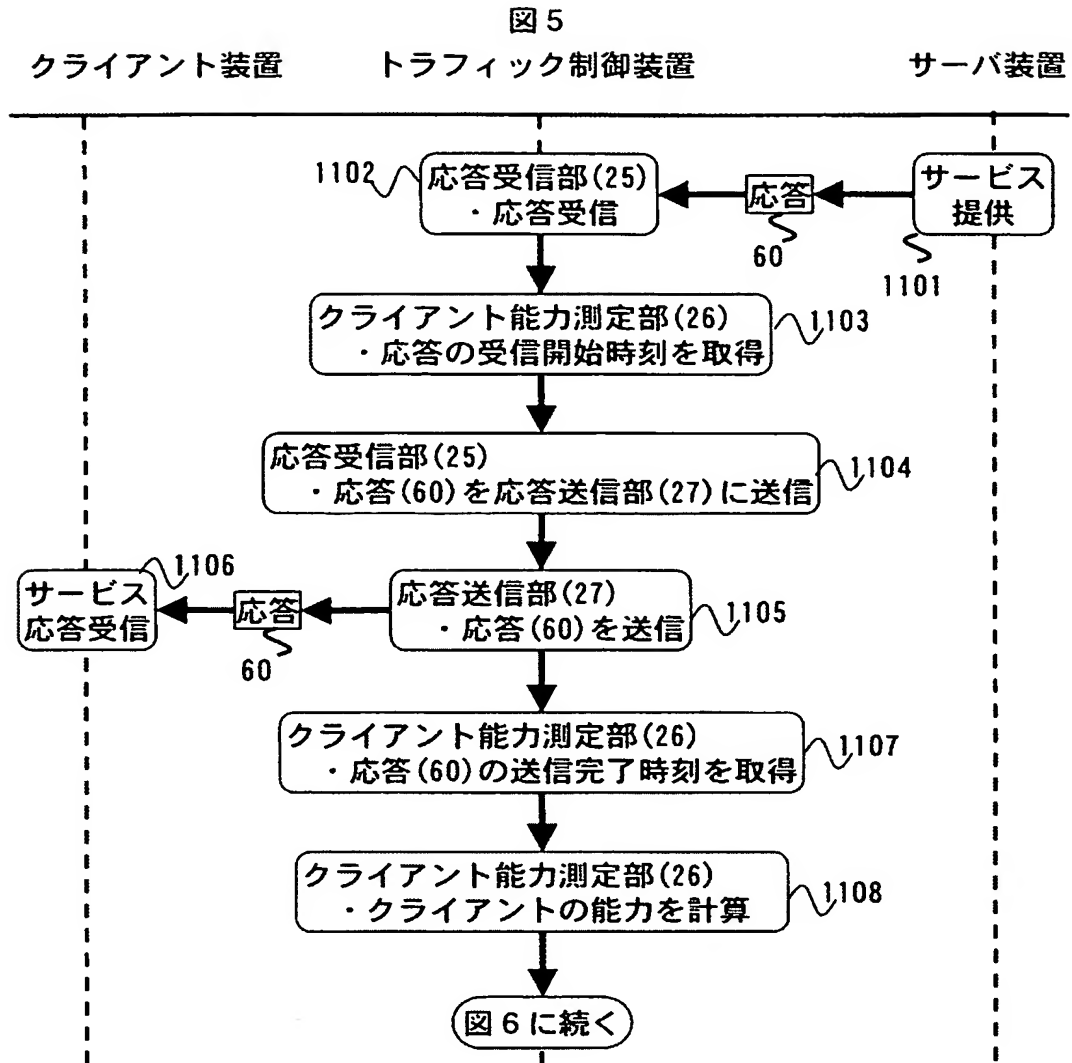


【図 4】

図 4

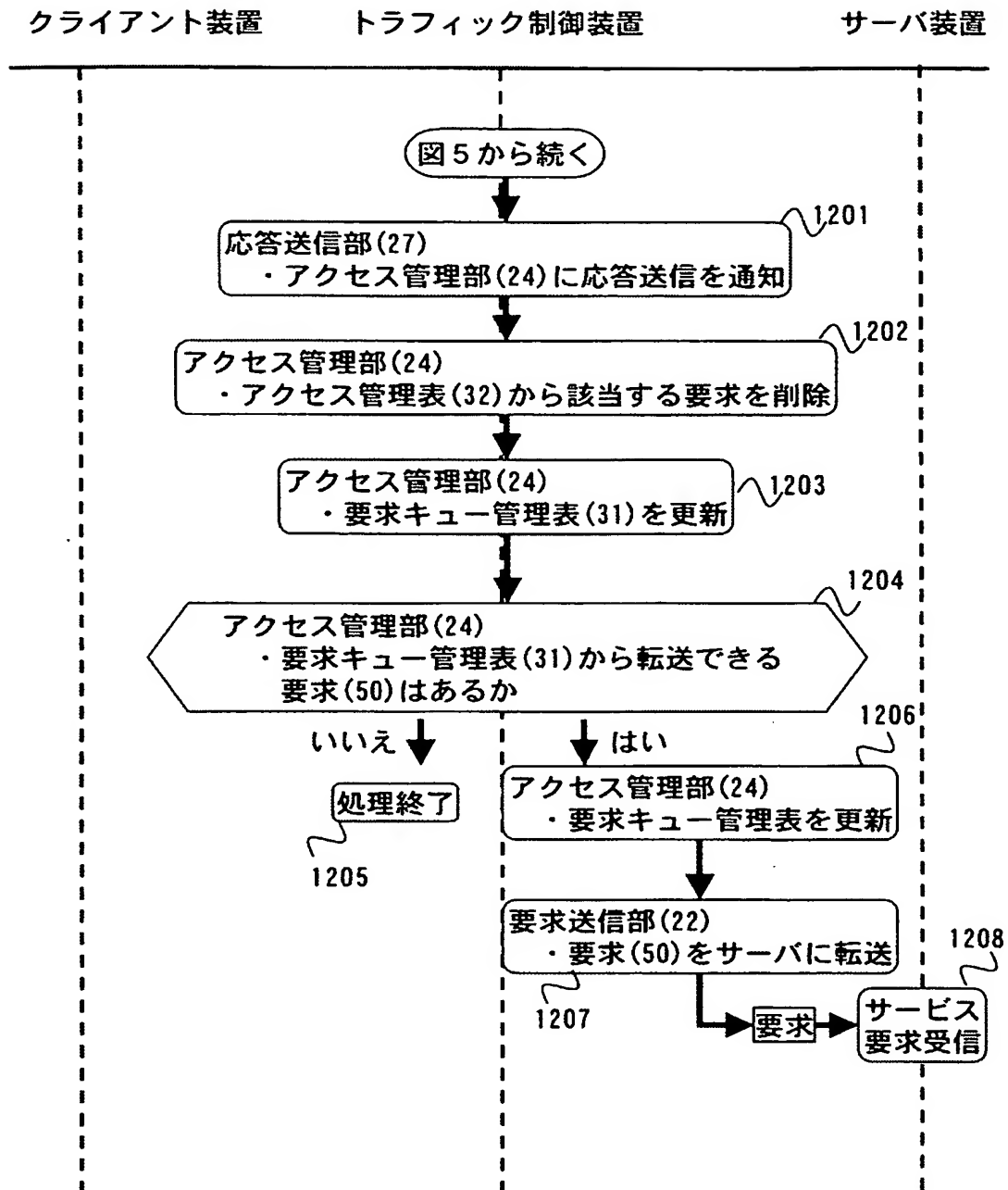


【図 5】



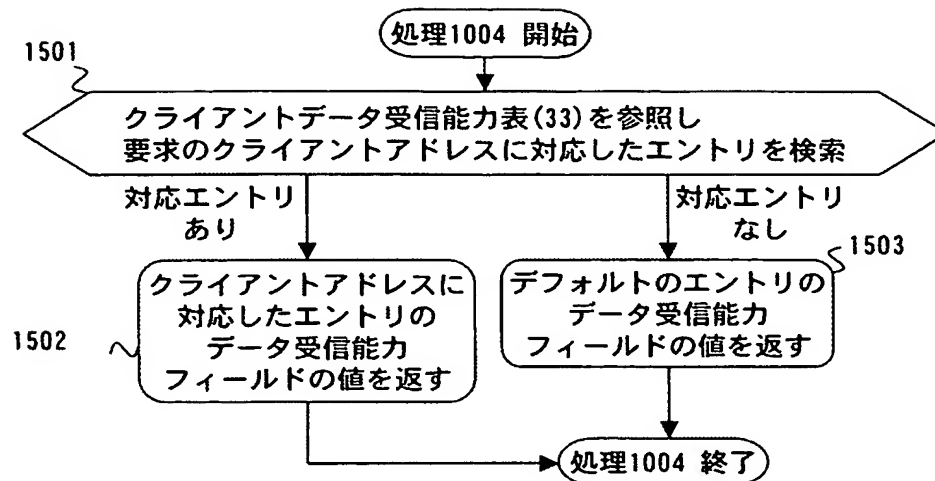
【図 6】

図 6



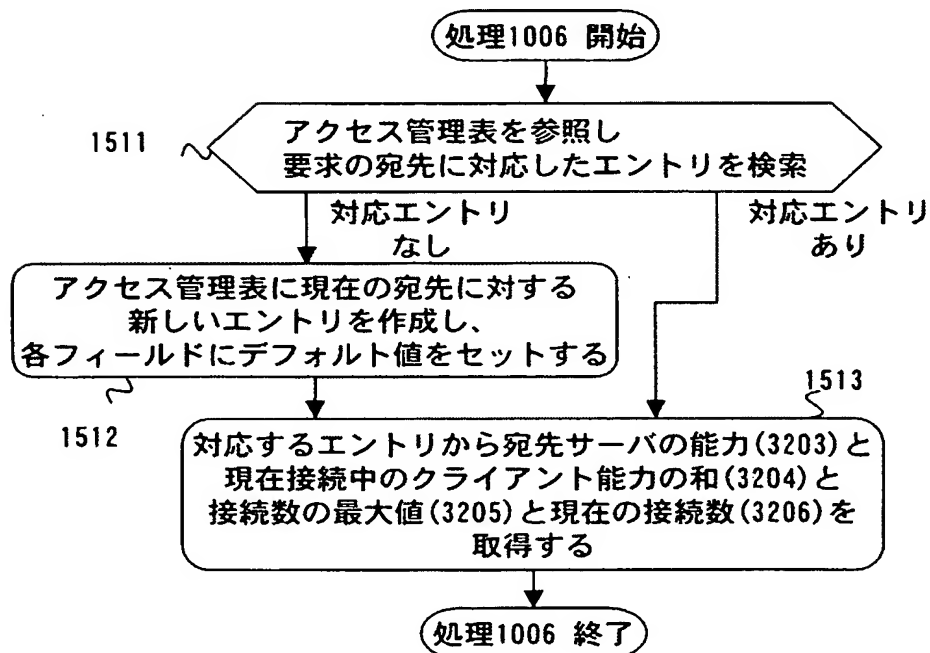
【図 7】

図 7



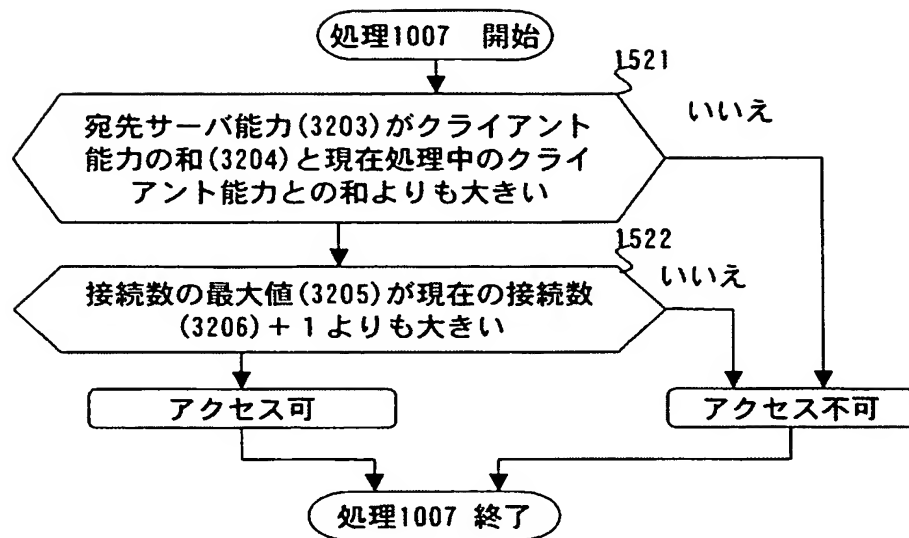
【図 8】

図 8



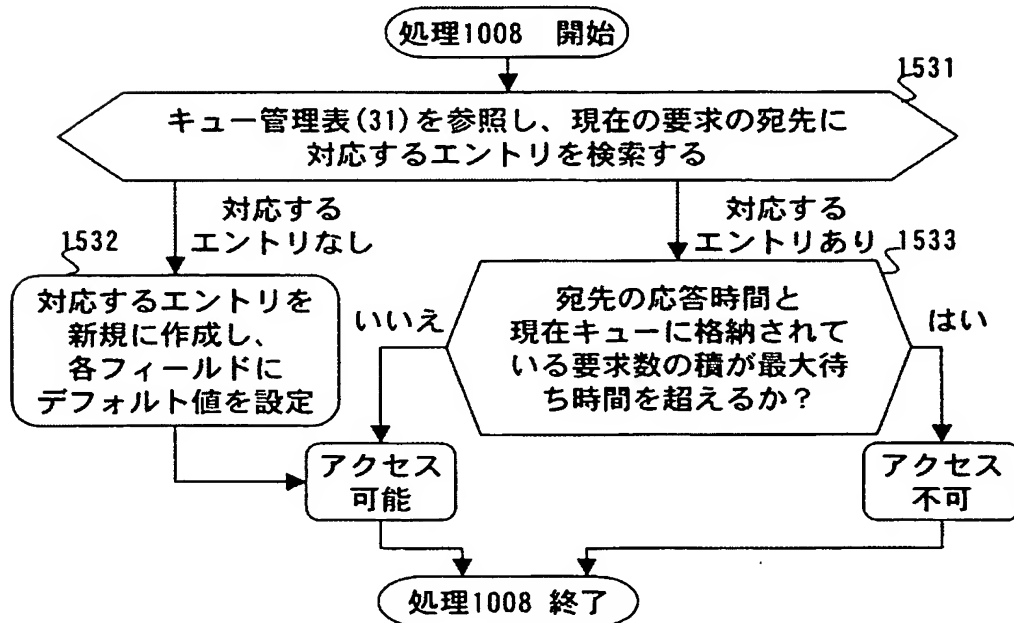
【図 9】

図 9



【図 10】

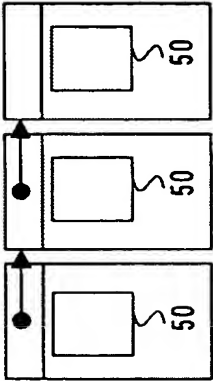
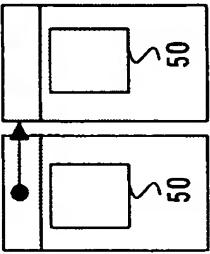
図 10



【図 11】

図 11

要求キュー管理表 (31)

宛先	キュー中の 要求数	最大 待ち 時間	応答時間			要求 (50) の リンクリスト
			平均	総和	処理数	
3101 ~	5	3.0	0.5	500	1000	
3101 ~	10	5.0	0.08	40	500	
.
.
3101 ~	0	3	0	0	0	
デフォルト						
Σ 3102	Σ 3103	Σ 3104	Σ 3105	Σ 3106	Σ 3107	Σ 3108

【図 12】

図 12

アクセス管理表(32)

	宛先	宛先サーバ 能力	クライアント 能力の和	最大接続数	現在の 接続数
3201 ~	10.20.1.1	100	40	100	30
3201 ~	10.30.2.2	120	70	200	155
	⋮				
3201 ~	デフォルト	100	0	100	0
	3202	3203	3204	3205	3206

【図 1 3】

図 1 3

クライアントデータ受信能力表 (33)

クライアント アドレス	クライアント データ 受信能力	送信開始 時刻	受信完了 時刻	データ サイズ
192.168.1.1	10	1058236564.00560612	1058236573.00279662	630000
172.16.10.20	20	1058236634.06005352	1058236643.02078422	500000
⋮				
デフォルト	10			

3301 ~

3301 ~

3301 ~

3302

3303

3304

3305

3306

【図 1 4】

図 1 4

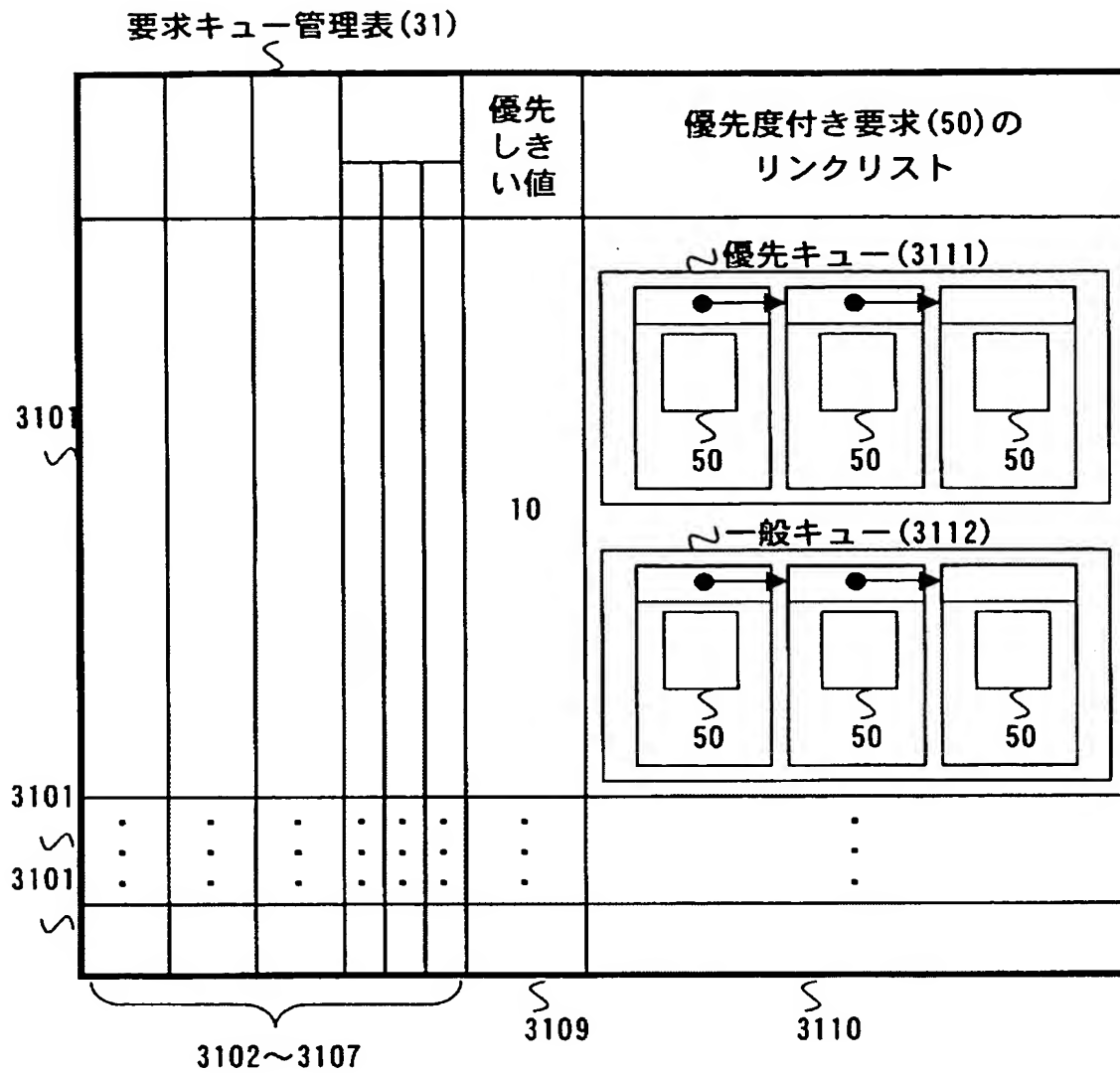
要求 (50)	宛先サーバアドレス	要求サービス名
---------	-----------	---------

51

52

【図 15】

図 15



【書類名】 要約書**【要約】****【課題】**

クライアントのネットワーク特性が多様な状況下で、サーバのスループットを維持し、かつサーバがダウンしないようにする。さらに、アクセス集中時に、クライアントが不確定な長時間サービス待ちの状態にならないようなサービスの提供を可能にする。

【解決手段】

クライアントからサーバへの要求を、トラフィック制御装置を経由して行い、トラフィック制御装置は、クライアントからの要求を制御する手段とクライアントのデータ受信能力を判定する手段とサーバに対する同時接続数を制御する手段を持つ。これにより、サーバのリソースを十分に利用できるように、かつ、サーバの性能以上に要求を転送しないように、サーバの同時接続数を制御する。

クライアントに対しては、要求のあったサービスの提供にある一定時間以上かかることが予想される場合には、要求を拒否し、ある一定時間以内にサービスが提供できることが予想された場合には、要求を受け付ける。

【選択図】 図 3

認定・付加情報

特許出願の番号	特願 2 0 0 3 - 4 1 8 9 0 5
受付番号	5 0 3 0 2 0 7 2 9 8 5
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 5 年 1 2 月 1 8 日

< 認定情報・付加情報 >

【提出日】	平成 15 年 12 月 17 日
-------	-------------------

特願 2 0 0 3 - 4 1 8 9 0 5

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 1 0 8]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所